DECISION SCIENCE CONSORTIUM, INC.

HOW ACCURATE ARE REAL WORLD FORECASTS AND ESTIMATES?

Robert C. Bromage

Decision Science Consortium, Inc.
7700 Leesburg Pike, Suite 421
Falls Church, Virginia 22043

September 1982

DTIC
ELECTE
OCT 1 1982

A

Technical Report No. 82-6

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO. AD A119809	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) HOW ACCURATE ARE REAL WORLD FORECASTS AND ESTIMATES?		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) R.C. Bromage		8. CONTRACT OR GRANT NUMBER(s) N00014-81-C-0330
9. PERFORMING ORGANIZATION NAME AND ADDRESS DECISION SCIENCE CONSORTIUM, INC. 7700 Leesburg Pike, Suite 421 Falls Church, Virginia 22043		10. PROGRAM ELEMENT. PROJECT. TASK AREA & WORK UNIT NUMBERS 61153N RR 014-11-01 NR 277-271
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Analysis Prog Mathematical and Information Sciences Division OFFICE OF NAVAL RESEARCH, Code 431 800 N. Quincy St., Arlington, Virginia 22217		12. REPORT DATE August 1982
		13. NUMBER OF PAGES 22
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office,		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for publication release: distribution unlimited. Reproduction in whole or in part is permitted for any purpose of the U.S. Government.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Forecasting, Estimation, Bayesian Statistics, Incoherence

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Modern forecasting and estimation techniques provide not only point estimates of unknown variables, but also associated intervals which reflect the expected accuracy of those estimates. Often different real world forecasts produce conflicting estimates and associated intervals of accuracy. This paper addresses the issue of how to make use of such estimates. It is argued that to both Classical and Bayesian statisticians the problem is essentially trivial. However, it is demonstrated that the assumptions required for a formal Bayesian approach are so sensitive to small changes, that the Bayesian approach has

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE
1 JAN 73
S/N 0102-014-6601 |

20. Abstract (continued)

dubious advantages over simple intuition. With the Classical attitude being
unhelpful in practice, it is argued that techniques should be developed which
combine formal Bayesian updating procedures with intuition. Two possible
techniques are explored. The first uses Bayesian updating with parameterized
likelihood functions. With suitable interpretation of the parameters, decision
makers can use their intuition to choose appropriate parameters. The second
technique allows for a number of alternate likelihood functions, combined prob-
abilistically according to the decision maker's judgment.

HOW ACCURATE ARE REAL WORLD FORECASTS AND ESTIMATES?

By

Robert C. Bromage

SUMMARY

Modern forecasting and estimation techniques provide not only point esti-
mates of unknown variables, but also associated intervals which reflect the
expected accuracy of those estimates. Often different real world forecasts
produce conflicting estimates and associated intervals of accuracy. This
paper addresses the issue of how to make use of such estimates. It is argued
that to both Classical and Bayesian statisticians the problem is essentially
trivial. However, it is demonstrated that the assumptions required for a
formal Bayesian approach are so sensitive to small changes that the Bayesian
approach has dubious advantages over simple intuition. With the Classical
attitude being unhelpful in practice, it is argued that techniques should be
developed which combine formal Bayesian updating procedures with intuition.
Two possible techniques are explored. The first uses Bayesian updating with
parameterized likelihood functions. With suitable interpretation of the
parameters, decision makers can use their intuition to choose appropriate
parameters. The second technique allows for a number of alternate likelihood
functions, combined probabilistically according to the decision maker's judg-
ment.

TABLE OF CONTENTS

1.0 INTRODUCTION

The need for the estimation of unknown quantities occurs daily in an extremely diverse field of decision making. Decisions from the mundane to those of national importance rest on assessments of how the weather will be tomorrow, how the economy will grow next year, or how far away is an enemy target. Assessments about future uncertain quantities are usually known as deductions or inferences. In either past, present or future, all are assessments of unknown quantities. Frequently these assessments are just "point" estimates: single values reflecting the most likely outcome. Thus, we may hear: "The forecast for tomorrow is for warmer weather," "The economy is predicted to grow by one percent next year," or "The target is estimated to be five hundred yards away." Use of the words forecast, predict, estimate, show that the quantity given is not precise. More informative assessments also provide a degree of confidence or "accuracy" in the point estimate. This may be a qualitative statement, such as "the temperature will be 65 degrees plus or minus five degrees," or the confidence may be related to the estimate as in: "plus or minus ten percent." Even more useful estimates will have either quantitative credible intervals or standard deviations provided with them, for example: "The temperature will be 65 degrees with a 95 percent chance of being within plus or minus ten degrees." Both Bayesian credible intervals and Classical confidence intervals are examples of such intervals. To avoid possible misinterpretation, the term "accuracy interval" will be used to describe the non-specific case.

Estimates, and associated accuracy intervals may be obtained in a variety of ways, ranging from direct judgment, intuition or guesswork, to highly sophisticated and complicated simulation models, and extremely carefully planned and constructed experimental trials. There is no situation where one and only one estimation procedure can be deemed the correct method. In theory, rough and ready procedures should have wide accuracy intervals, and increasingly accurate procedures should reduce those intervals. If one approach completely contains the information used in another, then the accuracy intervals produced by each ought to overlap considerably. For example, if one estimate of tomorrow's temperature is taken from a barometer reading

alone, whereas another uses both this reading, satellite pictures, and other such data, then although the two approaches may produce very different estimates, the way the extra information affects the estimates should be reflected in the changed accuracy intervals.

In practice, however, it is often the case that several different estimation techniques are, or could be, utilized, where the information used in one is not a subset of that used in another. Economic predictions can be obtained from a variety of models, each with its own set of assumptions about the way the economic world turns. For the submarine commander, a variety of different technologies are available to provide estimates of target range. Usually, where alternative techniques are available, it is known that no individual technique is exact, in that it fits the problem at hand exactly. Economic models are not exact. Experimental data is rarely directly applicable to the real world. However, the information obtained from models and experiments is considered useful enough to make the cost of modeling or experimentation worthwhile. Unfortunately, it often happens that different models will not only produce different point estimates, but also accuracy intervals which don't even overlap.

The problem to be faced is, how should the conflicting information from different sources be combined in an overall estimate? As French (1981) has observed, this problem is equivalent to that of the individual who assesses personal estimates of an unknown quantity in different ways, obtains different estimates, and has to "reconcile" them, i.e. the reconciliation of incoherent judgments (Lindley, Tversky and Brown, 1979). In this paper, we examine ways in which the problem has been approached, and suggest ways in which conflicting information might be used constructively in estimation procedures.

2.0 FORMAL STATISTICAL CONSIDERATIONS

Given that the problem as described is so prevalent in the world, one might
think that it would have received widespread treatment by statisticians and
people directly concerned with the problem as they are directly affected by
it. This is not the case, though, primarily because both schools of statis-
tical thought, Classical and Bayesian, are formulated in such a way that for
either school the problem is theoretically trivial (as explained below), and
so unworthy of consideration. Such a stance is not particularly useful for
people who face the problem in the real world, and who generally use a variety
of ad hoc techniques to aid their judgment.

2.1 The Classical Statistical Approach

A good critique of the classical statistical approach can be found in Bunn
(1978), in which he suggests that the attempt to follow the "scientific
method" is the underlying philosophy behind the approach. This method simply
states that the best estimation technique should be chosen, to the exclusion
of all others, with certain cost considerations. Different methods are com-
pared by how well they may have performed on past data. As a result, dif-
ferent approaches are explicitly developed to perform well on past data, by
using the past data to develop and estimate parameters in the models. At
one extreme, there is the area of time series analysis, which is developed
using only past data. At the other extreme, sophisticated simulation models
are developed using seemingly natural relationships (e.g., linear) between
understandable parameters, but still the constants in the equations are
developed to fit past data as well as possible. The problem then becomes
one of applying Ockam's Razor, for with enough parameters, an arbitrarily good
fit can be obtained.

When it comes to experimentation, combinations of experiments can be consi-
dered. A classical confidence interval for a variable quantity is a range
of acceptable values for that quantity. The process is to assume a hypothetical
value for the quantity and to assess the probability of obtaining the observed
data, or more "extreme" data. The hypothesis is rejected if this probability

is below some arbitrarily stated value (known as the size). Two problems can arise when combinations of experiments are considered. First, it may not be clear how the experiments should be combined. If the data cannot be lumped together as if it were one large experiment, it may be difficult to assess which data are more "extreme" than others. Second, it may turn out that with the combined data, no acceptable hypotheses are available. More generally, as the experiments provide more inconsistent data, the intersection of the two confidence intervals becomes smaller. This has meant that data from different experiments are usually only combined if it is felt that the data are "identical." If this is not the case, one or both experiments are rejected.

Assumptions have to be made about any model, test, or forecast. The classical statistical philosophy, having accepted an approach, declares that all such assumptions are perfectly held. Alternatively, if an assumption is rejected, then so are all the results. Sensitivity analyses may be performed on particular values, but there is no formal method for incorporating the results of such analyses in the estimate. Assumptions of model structure are inviolate. One either accepts the model completely, or rejects its results. There can be no compromise.

2.2 Ad Hoc Procedures

Aware of the inadequacies of the classical statistical philosophy in particular areas, researchers, and particularly psychologists who feel comfortable without an underlying formal mathematical rock, have resorted to examining ad hoc techniques for resolving the problem. By "ad hoc," we mean simply that the basis for the technique has intuitive rather than formal rationale. Such efforts have been applied particularly where the unknown quantity is a "probability," and the conflicting estimates are expert assessments. A review of this literature, together with a discussion of experimental results, can be found in Seaver (1976). Generally, the intuitive rationale has been that the estimates should be combined using a linear combination of them. Assessment of the weights in such a linear combination has been treated in several ways, from the simple to the sophisticated. In the specific case of

probability assessment, the reader is referred to Stone (1961), Roberts (1965), and De Groot (1974). Bunn (1978) applies two separate techniques in the synthesis of forecasting models. Non-linear procedures are well reviewed by Seaver, for probabilities, and generally have not been considered for other variables. The problem with ad hoc procedures is that they can only be justified empirically, and not formally. It can therefore become difficult to justify their use in areas where empirical testing of different approaches has not been adequately carried out, which adequacy is very much a subjective assessment.

2.3 The Bayesian Approach

According to Bayesian statisticians, the problem is just as trivial as it is to classical statisticians. For "all" a decision maker has to do is specify the relevant priors and likelihood functions, and there is the posterior distribution for the variable under consideration, complete with variances, confidence intervals and so on.

Morris (1974, 1977) pioneered the Bayesian approach to combining estimates, with particular accent on expert judgments. Winkler (1981), provides a Normal model that incorporates dependencies between the assessments. Although the accent is on expert judgments, he recognizes the wider area of applicability.

Lindley (1981) has developed an almost identical model applied to estimates of variables other than probability, specifically estimates of range, based on various different ranging techniques. While Winkler's paper concentrates primarily on the problem of dependence between estimates, Lindley provides more depth to the assumptions underlying the model.

Although the Bayesian approach is formally correct, without making assumptions such as those of Winkler and Lindley, it requires a lot of hard work with difficult judgments. Not only are priors and likelihood functions often difficult to assess, but even with Normal approximations, assessments of correlation are notoriously difficult. Kadane et al. (1980) have proposed tech-

niques for facilitating the assessments of correlation coefficients, but as yet validation of these techniques has not been performed.

Lindley, Tversky and Brown (1979) formulated two Bayesian approaches for combining one individual's probability estimates, arrived at from different perspectives, and French (1980) modifies one of these approaches to address this problem. Both these studies assume that Bayesian updating is the correct way to combine probabilities, and concentrate on the interpretation of the likelihood functions and priors.

3.0 PROBLEMS OF RECONCILIATION

To the Bayesian statistician, the formal problem of combining probability
distributions might appear to have been solved. If θ is the variable under
consideration, and $\underline{x}$ is a vector describing all the distributions, then in
theory, one obtains the posterior for θ using Bayes' formula:

$$f(\theta|\underline{x}) \propto f(\underline{x}|\theta)\pi(\theta)$$

where $f(\theta|\underline{x})$ is the posterior distribution for θ, $f(\underline{x}|\theta)$ is the likelihood
function for $\underline{x}$, and $\pi(\theta)$ is the prior distribution. The only remaining prob-
lems are operational - how to apply this in practice, and it would appear
that Winkler and Lindley have provided major inroads into this problem.

However, it is at this point that the fundamental question of coherence
raises its head: why should the use of Bayes' formula be better than a
direct assessment of the posterior? Although Brown and Lindley (1982) have
addressed it, many Bayesian statisticians still feel that this question is
a non-question. To them, the Bayes formula is by definition the correct
way to analyze the problem. However, as we shall show, recent attempts to
apply the Bayesian approach in the practice of combining distributions have
brought out the full colors of the coherence problem.

Traditionally, Bayesian analysis tends to have been directed at simple cases
where, for example, likelihood functions are easily assessed, and independence
between component likelihoods is assured. Often the variable under considera-
tion is the parameter of the likelihood function in a classical statistical
sense, so that the likelihood function is unequivocally defined. It is clear
that one's posterior feelings about a variable will depend, in a causal sense,
on one's prior feelings, and so, as prior feelings have to be considered any-
way, Bayes' formula will make formally correct use of the data. This argument
can be found in Phillips (1973) and we agree with it. However, as we now show,
the assumption that the same argument holds when the likelihood function is
not well defined is very dubious.

3.1 A Practical Application

In a recent project, the Bayesian techniques developed have been applied
to a ranging problem. In this problem, measures of the range of an object
are obtained using various different techniques, from sophisticated scienti-
fic devices, to human judgment. All the techniques produce both estimates
of range and of confidence intervals or variance - some measure of accuracy
of those estimates. Note that the likelihood functions in such a case are
not perfectly defined. Cohen and Brown (1980) considered the use of inde-
pendent Normal models of likelihood (and flat priors). A direct consequence
of this was that, no matter how wide the range estimates were, for given
variance estimates, the combined distribution always had the same variance,
which was smaller than the smallest given variance estimate. On the other
hand, using independent t-models, Lindley (1981) demonstrates that as range
estimates diverge, the variance of the combined distribution also diverges,
owing to the nature of the "poly-t" distribution (Drèze, 1971) that results.
For two estimates, with equal variances (for simplicity), the differences
between the resulting posteriors are depicted in figure 1 (for large dif-
ferences in estimates):

Normal Model:

←— as estimates diverge —→

t-Model:
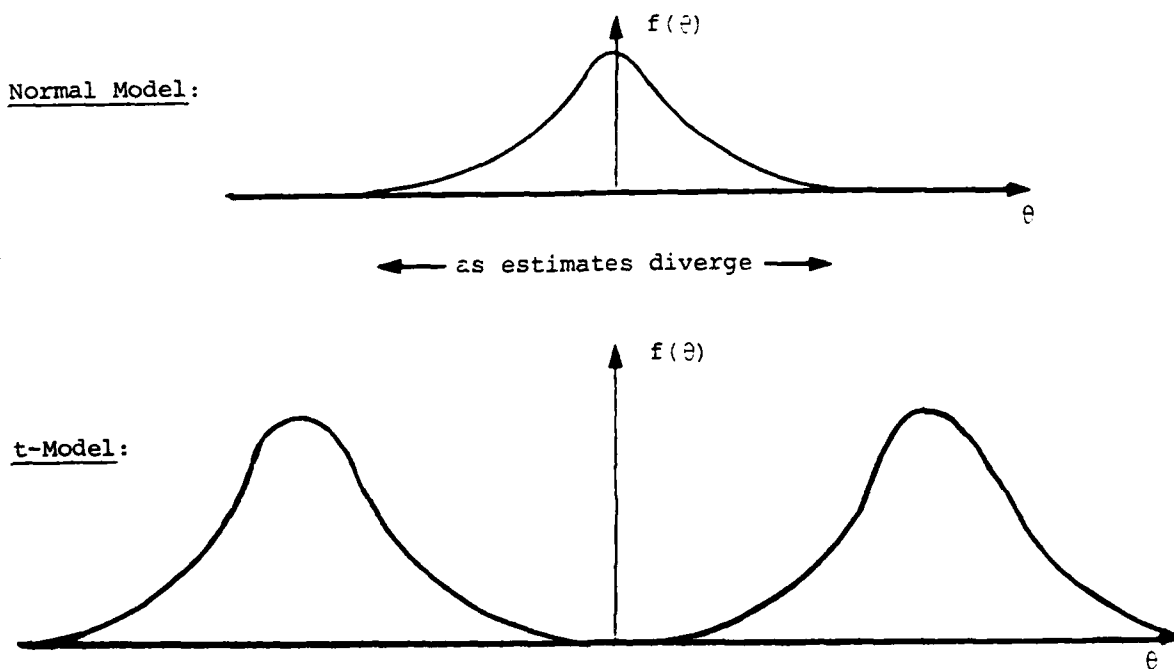
Now it must be remembered that the likelihood distributions assumed in each case have the same means, almost the same variances (which could be justifiably made the same in a manner Lindley proposes), and are uni-modal distributions. It is, in our view, inconceivable that a subject could, in making a subjective assessment between the likelihood distributions, claim to appreciate the difference between two such similar distributions. Yet the posteriors obtained are wildly different.

Formally, the Bayesian statistician now finds himself in a classical position. He must choose the model whose assumptions seem most reasonable (presumably the t-model in this case, being less restrictive), and use it to the exclusion of all others. In reality, it is far more tempting to compare the models on the basis of their effects on the posterior distribution. In this case, for example, we might argue for the t-distribution on the grounds that it seems intuitively better that the posterior variance should increase as the estimates diverge. However, such an argument is formally unacceptable, because it uses direct assessments of the posterior to choose the likelihoods - exactly the opposite of what Bayes' theorem suggests should be done. The reasoning is circular. However, it is not very comforting to know that the posterior distribution obtained using Bayesian updating is highly sensitive to a model assumption that is almost certainly wrong. The coherence question rises strongly. Why should we put blind faith in a model whose assumptions do not hold, when the results contradict what we see with our own eyes? Why should one be the illusion any more than the other?

3.2 A Frequentist Experiment Analogy

It often helps to understand strange statistical happenings by consideration of an experimental analogy. In this experiment we have a bag full of different colored balls, and we wish to determine the proportions of each color by repeated drawing from the bag. Unfortunately someone else does the drawing for us, and we can only collect data on each 100 drawings (with replacement). After two lots of 100 drawings, the first lot produces p ($\geq$ 50) blue balls and (100 - p) red balls, the second produces (100 - p) blue balls and

p red balls. Now, if all the experimental assumptions of randomness are made, and with a uniform prior on p, the proportion of blue balls, the posterior distribution for θ will have a mean of .5 and a standard deviation of .035, (taken from its beta distribution), and this is __independent__ of p. Furthermore, given the experimental assumptions, the Normal approximation is a good one for the final distribution. So we see that picking Normal likelihoods is equivalent to having complete faith in the experimental assumptions.

Suppose now that p actually turned out to be 80. With a uniform prior, one set of data provides a mean for θ of .79, and standard deviation .04. Using the other set, the mean for θ is .21 and standard deviation again .04. The sets of data clearly conflict, as even the 99% credible intervals do not overlap. In practice, such data would probably shake our faith in the validity of the experiment. How we would interpret the data would depend upon how easily our faith was shaken, and what alternative models of the process might be considered. The example shows, though, that the data will not only change one's assessments of the variable, but also of the validity of the models underlying the relation between variable and data. In this example, the data might suggest that one of the sets of data was reported wrongly. However, to incorporate such data in the likelihood functions would require consideration of all other strange pairs of outcomes, and it would be totally impractical to try and pre-guess every strange outcome and thus assess probabilities for each model. Furthermore, the flash of inspiration that particular sets of data produce is never pre-guessed, so that that process may not only be impractical, but also impossible. Having said that, a limited assessment of different models, rather like Bunn (1978) proposed, but applied to likelihood models as opposed to posteriors, might be fruitful. This will be explored in section 3.3. Finally, we might observe that assessing t-likelihoods seems to fit this anomaly quite nicely, for as the estimates get closer, the t-distribution puts an increasing weight on values between the estimates, thus seeming to suggest that the usual model has more weight than the alternative. Such intuitive notions will be addressed later.

3.3 A Model Conditioned Approach

In many real world cases, modeling techniques, experimental data from artificial trials, or past data are known not to be absolutely appropriate for the problem in question, but they still may be useful, and possibly the best information available. The weighting procedures described in section 2.2 consider the case where each model produces a distribution for the variable in question, and, given model acceptability, it is assumed that the model distribution will be the distribution to be used. Formally the reasoning is:

$$f(\theta|\underline{x}) = \sum_i f(\theta|\underline{x} \& M_i)p(M_i|\underline{x}) \qquad \text{when } M_i \text{ is the } i^{th} \text{ model,}$$

$$\text{But} \quad f(\theta|\underline{x} \& M_i) = f_{M_i}(\theta) \qquad \text{by the assumption of model use.}$$

$$\longrightarrow f(\theta|\underline{x}) = \sum f_{M_i}(\theta)p(M_i|\underline{x}).$$

A similar analysis can be attempted for the case where the models are not necessarily exclusive. This type of analysis works well where, for example, the models use expert judgments. In such cases, Bayesian updating is inappropriate, because the likelihood functions are difficult to comprehend. In terms of expert judgments, it would involve assessing the probability that an expert would provide a probability of p, given the occurrence of an event. Thus we would argue that the "ad hoc" procedures of section 2.2 do indeed have a subjective (if not Bayesian) justification.

In other cases, however, the models, or some of the models, provide likelihoods rather than direct assessments of the variable distributions. The ball and bag model of the previous section is an example. Trial data for instrument calibration may be another. The validity of the model could be modeled in terms of a number of discrete possibilities, or of a continuum of possibilities. Here we consider just the case where either the model is valid, or it is not.

In this case, we have:

$$f(\theta|x) = f(\theta \ \& \ M \ x) + f(\theta \ \& \ {\sim}M|x)$$

where M is supposition that the model is valid, and ${\sim}M$ that it is not.

Then $\qquad f(\theta|x) = f(\theta|x \ \& \ M)p(M|x) + f(\theta|x \ \& \ {\sim}M)p({\sim}M|x).$

$\longrightarrow \qquad f(\theta|x) = \left[\dfrac{f(x|\theta \ \& \ M)^-(\theta|M)p(M|x)}{\int_\theta f(x|\theta \ \& \ M)^-(\theta \ M)d\theta}\right] + \left[f(\theta \ x \ \& \ {\sim}M)p({\sim}M|x)\right] \quad (1)$

The first bracketed term on the right hand side of equation (1) is the usual
Bayesian updating formula, weighted by $p(M|x)$, the probability that the model
is valid, given the data. The second bracketed term may also be formulated
in the same way, if it is appropriate. Alternatively, it may be that
$f(\theta|x \ \& \ {\sim}M)$ is independent of x, i.e., $f(\theta|x \ \& \ {\sim}M) = {}^-(\theta|{\sim}M)$. The way it is
treated will depend very much on the individual problem. It may be felt
that this approach appears too complicated to be practical, but we hope to
demonstrate that it forms the basis for useful practical techniques.

4.0 TWO APPROACHES TO PRACTICAL SOLUTION

In section 2, two problems were associated with the Bayesian updating approach. The first was that small differences in likelihood functions can produce large changes in posterior distributions. This means that making assumptions about the form of a likelihood function can produce posterior distributions that are counter-intuitive. Second, likelihood functions can sometimes be conceptually difficult to comprehend. In the second case, we suggested that alternative procedures should receive more formal acceptance, but it is the first case that is pursued here.

Given that strict adherence to Bayesian updating may lead to large errors in posterior distributions, it is the philosophy of the advocates of exploiting incoherence (Brown & Lindley, 1982), that consideration of both direct intuitive judgment and of implied Bayesian analysis should be made. The question then becomes, what is the best mix of the alternate approaches? In this section we propose two possible answers to that question, each of which may be appropriate in certain situations.

4.1 Recap of Suggested Likelihood Functions

Consider again the two forms of likelihood function that have been suggested in the literature, namely Normal and t. In this recap, we shall restrict attention to the case where there are two estimates, each with the same variance, s^2, and means at $\pm m$. The prior is assumed to be uniform.

Normal. The combination of two normal distributions with these means and variance produces another normal distribution, with mean 0, and variance $\frac{1}{2}s^2$. If the variances are different, the combined variance will be smaller than the smaller of the two likelihood variances. The combined distribution is independent of the distance between the means of the two likelihoods.

Student's t. The distribution produced by the combination of two t-distributions is dependent on how far apart (relative to the standard deviation) the means are. Lindley (1981) shows that for the likelihoods described,

the distribution produced will be bi-modal if:

$$m^2 > s^2 \nu.$$

where ν is the number of degrees of freedom in each t-distribution. A
little algebra also shows that as $m \to \infty$, for fixed s and ν, the distribu-
tion tends toward two disjoint distributions each with a probability of $\frac{1}{2}$
of occurring, situated at $+m$ and $-m$, and with variances of $s^2\nu/(\nu-2)$ (if $\nu>2$).
So although the overall variance (about the mean of 0) is large, the distri-
bution is nevertheless confined to two narrow areas.

Now, for given ν, the resultant posterior from using the t-distribution may
not seem particularly intuitive, but the advantage of the distribution is
that by varying the parameter, the posterior will change, whereas with the
Normal distribution, there is no flexibility.

4.2 The Use of Parameterized Likelihoods

The choice of parameter ν has been given the usual interpretation by
Winkler (1981) and Lindley (1981), but this interpretation is not entirely
sound, because in neither case is uncertainty about the variance related to
trials, or if it is (in the ranging case), the number of trials is so large
that a Normal distribution may as well be used. A better interpretation of
ν is that, in recognition of the fact that the experimental data do not fit
the real situation exactly, the choice of ν reflects intuition about the
relation between experiment and real world. Of course this interpretation
is not exact, but it may be the basis for a link between intuition and
Bayesian updating.

Two points emerge from this discussion. First, in order for the use of
parameters to be effective, the decision maker must have a good notion of
what the parameter means, and thus how it might be varied according to his
or her intuitions.

The second point is that the t-distribution may not provide desirable posteriors for any value of ν. In which case, are there other parameterized distributions which might fit a particular problem better than the t-distribution? Ideally, what we would like is a distribution with a small number of parameters which can be varied to fit a wide possible number of problems. However, we must also beware of hoping that the distribution chosen will fit every case perfectly.

This observation provides a basis for suggesting when using the t-distribution will be acceptable. For it may be that in many situations what would ideally be desired was a unimodal distribution whose variance increased in a certain manner as the two estimates diverge. The t-distribution may suit this case so long as the estimates do not diverge too much. How much is too much? This must be subjectively estimated, for a bi-modal shape may be quite acceptable as long as there is sufficient weight in the centre. Pictionally we may have three possible t-distributions, 2 acceptable and one not:

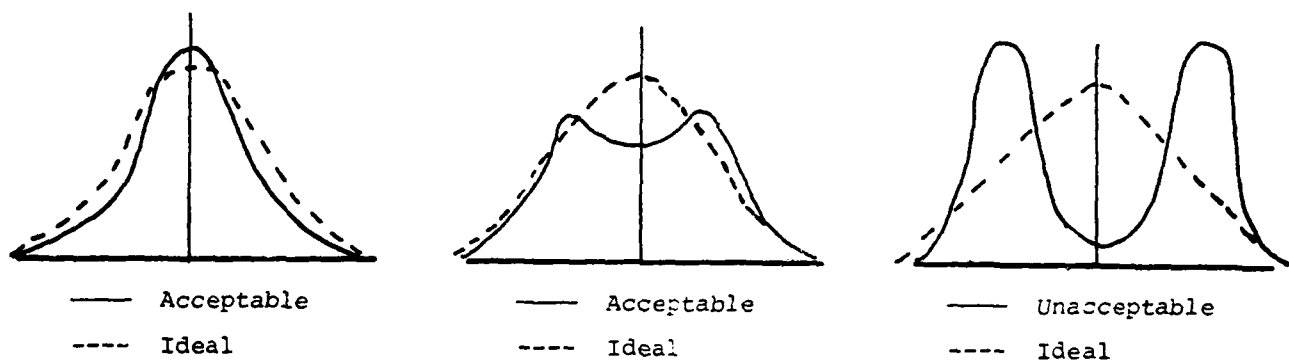|		
—— Acceptable	—— Acceptable	—— Unacceptable
---- Ideal	---- Ideal	---- Ideal

Figure 2

The acceptability decision rule may be that the height of the posterior at the mean must be at least, say, half that of the biggest node. Using this decision rule, it can be shown that this means that m^2/s^2 is less than about $2\nu(2^\nu - 1)$, so, for example, if $\nu = 5$, the ratio of mean to standard deviation of assessments must be less than about 17.

-15-

There are, of course, several alternatives to using the t-distribution.
Suppose that we desire a posterior distribution which is always unimodal,
but whose variance increases as estimates diverge. If the likelihoods are
of the form $f(x - \theta)$, then the posterior is proportional to $f(x-\theta)f(x+\theta)$.
For the posterior to be unimodal but with an increasing variance, we require
first that $f[(x-\theta)]f[(x+\theta)]$ be a maximum when $\theta = 0$, with no local minima,
but that $\int \theta^2 f(x-\theta)f(x+\theta)d\theta$ should increase in x. The first condition re-
quires that $f''(x)f(x) - [f'(x)]^2$ be negative for large positive x. Appendix
A shows that this is never true for inverse polynomial functions, like the
t distribution, but is true for exponential functions of the form $e^{-\theta^y}$ where
y is greater than 1. The function $e^{-\theta}$ is of special interest, because it
provides a zero value for the differential equation for all values of t
between -x and x. Hence, if two likelihood functions have this form, and
the same variance, the posterior will have the appearance of Figure 3:
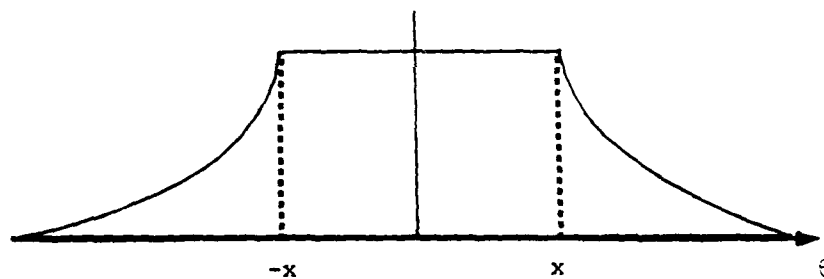

Figure 3

What this means is that the posterior will behave in the desired manner if
the tails of the likelihood functions look something like $e^{-\theta}$. As we know
that likelihood functions of the form $e^{-\theta^2}$ result in a posterior that is in-
variant with respect to the estimate difference, we can safely predict that
functions of the form $e^{-|\theta|^y}$ where $1 \leq y < 2$ will give likelihoods that will
produce diverging posteriors with values of y near 1 causing more divergence
than those near 2. The problem with such distributions is that they are
difficult to handle analytically, so that it is difficult to place an inter-
pretation on y, and the use of the modulus makes computation of the posterior
complicated.

The problem of intractability of posteriors is very acute, because so few likelihoods are both sufficiently general purpose, and at the same time possess this property. It seems inevitable that the use of likelihoods with numerical analyses of posteriors will need to be performed, such as, indeed, the t-distribution. Once this is accepted an almost limitless number of distributions are available, including in particular, beta, inverse-beta, and gamma distributions. The computational complexity of the resultant posteriors is such that the analysis could only feasibly be performed on a computer. However, based on the analysis of the t-distribution, we would hypothesize that while other single parameter distributions might outperform the t-distribution in certain ways, in order to provide more flexibility, it is likely that double parameter distributions will be necessary. Also, based again on observations for the t-distribution, we would predict that such a distribution would provide a sufficiently rich set of posteriors that it is likely to be unnecessary to move to triple parameter distributions.

4.3 The Use of Probabilistically Combined Likelihoods

Consider again formula (1) of section 3.3:

$$f(\theta|x) = \frac{f(x|\theta \ \& \ M)\pi(\theta|M)}{\int_\theta f(x|\theta \ \& \ M)\pi(\theta|M)d\theta} \ p(M|x) + f(\theta|x \ \& \ \sim M)p(\sim M|x).$$

For the case where there are several identifiable mdoels, each of which may be updated in Bayesian fashion, we obtain:

$$f(\theta|x) = \sum_i \left[\frac{f(x|\theta \ \& \ M_i)\pi(\theta|M_i)}{\int_\theta f(x|\theta \ \& \ M_i)\pi(\theta|M_i)d\theta} \right] \ p(M_i|x).$$

Within the summation sign, this formula contains two parts. The bracketed part is the Bayesian updating formula given acceptance of a particular model. The remaining part is the probability of the model being correct,

<u>given</u> the data that has been observed.

To the Bayesian statistician, the formally correct approach might be to obtain $p(M_i|x)$ from Bayes formula:

$$p(M_i|x) = \frac{p(x|M_i)\pi(M_i)}{\sum_j p(x|m_j)\tau(M_j)}$$

which is appealing because the likelihood functions $p(x|M_j)$ will usually be of "classical" type. However, such an approach denies the essentially recursive nature of model/hypothesis generation and testing-the argument of Section 3.2. In strictly Bayesian terms, that argument implies that prior distributions for models require an effort that is beyond human capability. On the other hand, the posterior distribution, $p(M_i|x)$, the probability of model acceptability given our current state of knowledge, is a very understandable uncertainty, which allows for the generation of previously unthought of models.

5.0 COMPARISON OF THE TWO TECHNIQUES

Two different techniques have been presented which are designed to allow
decision makers to combine conflicting information from different sources
in forecasting or estimation. A common feature of these techniques is that
they employ the decision maker's intuition as well as formal properties of
probability theory. The appeal of one or the other of them is therefore
liable to be context-dependent.

There is, however, one distinguishable difference between the approaches. The
model conditioned approach assumes that only one (if any) model, is "correct,"
but that the correct model's identity is unknown. The parameterized likeli-
hood approach, on the other hand, assumes that each model's output is useful,
at least to a certain extent. This difference probably will favor the para-
meterized likelihood approach for decision makers, but must be weighted against
the fact that that approach is less easily understood--the meaning and use of
the parameters have to be explained. Finally, it should be remembered that
the purpose behind providing more than just point estimates of unknown quanti-
ties is to provide people who would use those estimates with suggestions of
literally how much confidence they can place in those estimates. The combina-
tion of judgment and probabilistic analysis allows decision makers to explore
their assumptions about the validity or applicability of the assumptions that
must be made in using models, experimental or past data, and so on, in real
world decision making.

Appendix A

Let $f(x)$ be a probability density function of the form $f(x) \times (p(x))^{-m}$, where $p(x)$ is a polynomial in x, and $m > 0$. Then $f''(x)f(x) - [f'(x)]^2 > 0$ for x sufficiently large.

Verification

$$f(x) = k(p(x))^{-m}$$

$$\longrightarrow \quad f'(x) = -km(p(x))^{-(m+1)}(p'(x))$$

$$\longrightarrow \quad f''(x) = km(m+1)(p(x))^{-(m+2)}(p'(x))^2 - km\,p(x)^{-(m+1)}p''(x)$$

$$\longrightarrow \quad f''(x)f(x) - [f'(x)]^2 = k^2 m \left\{ (m+1)\left\{p(x)\right\}^{-2m-1}(p'(x))^2 - \right.$$
$$\left. (p(x))^{-(2m+1)}p''(x) - m(p(x))^{-2(m+1)}(p'(x))^2 \right\}$$

$$= k^2 m \left\{ (p(x))^{-2(m+1)}\,p'(x))^2 - (p(x))^{-(2m+1)}p''(x) \right\}$$

$$= \frac{k^2 m}{p(x)^{-2(m+1)}}\left((p'(x))^2 - p(x)p''(x) \right).$$

Hence

$$f''(x)f(x) - [f'(x)]^2 > 0 \longleftrightarrow (p'(x))^2 - p(x)p''(x) > 0$$

$$\text{Let } p(x) = ax^n + bx^{n-1} + \ldots$$

$$\text{Then } (p'(x))^2 = a^2 n^2 x^{2(n-1)} + 2abn(n-1)x^{2n-3} - \ldots$$

$$\text{and } p(x)p''(x) = a^2 n(n-1)x^{2(n-1)} + 2abn(n-1)x^{2n-3} + \ldots$$

As $x \longrightarrow \infty$, $(p'(x))^2 - p(x)p''(x) > 0 \longleftrightarrow$ highest term in $x > 0$

Term in $x^{2(n-1)} = a^2 n^2 - a^2 n(n-1) = a^2 n > 0$

Hence $(p'(n))^2 - p(x)p''(x) > 0$ and $f''(x)f(x) - [f'(x)]^2 > 0$ as $x \rightarrow \infty$

-20-

REFERENCES

Brown, R.V., and Lindley, D.V. Improving judgment by reconciling incoherence. Theory and Decision, 1982, Vol. 14, Issue 2.

Bunn, D.W. The synthesis of forecasting models in decision analysis. Basel: Birkhauser Verlag, 1978.

Cohen, M.S., and Brown, R.V. Decision support for attack submarine commanders. Falls Church, VA: Decision Science Consortium, Inc., October 1980.

DeGroot, M. Reaching a consensus. Journal of the American Statistical Association, 1965, 60, 50-62.

Dreze, J.H. Bayesian regression analysis using ploy-t densities. CORE Discussion Paper 7613, Universite Catholique de Louvain, Heverlee, Belgium, 1976.

French, S. Consensus of opinion. European Journal of Operational Research, 1981, 7(4), 332-340.

French, S. Updating of belief in the light of someone else's opinion. Journal of the Royal Statistical Society, Series A, 1980.

Kadane, J.B., Dickey, J.M., Winkler, R.L., Smith, W.S., and Peters, S.C. Interactive elicitation of opinion for a normal linear model. Journal of the American Statistical Association, 1980, 75(372), 845-854.

Lindley, D.V. Reconciling continuous probability assessments. Submitted to Operations Research (1981).

Lindley, D.V., Tversky, A., and Brown, R.V. On the reconciliation of probability assessments. Journal of the Royal Statistical Society, Series A, 1979, 142, 146-180.

Morris, P. Combining expert judgments: A Bayesian approach. Management Science, 1977, 23, 679-693.

Morris, P. Decision analysis expert use. Management Science, 1974, 21, 1233-1241.

Phillips, L.D. Bayesian statistics for social scientists. New York: Thomas Y. Crowell Co., 1973.

Roberts, H. Probabilistic prediction. Journal of the American Statistical Association, 1965, 60, 50-62.

Seaver, D.A. Assessment of group preferences and group uncertainty for decision making (SSRI Research Report 76-4). Los Angeles, CA: University of Southern California, Social Science Research Institute, 1976.

Stone, M. The opinion pool. Annals of mathematical statistics, 1961, 32, 1339-1342.

Winkler, R.L. Combining probability distributions from dependent information sources. Management Science, 1981, 27(4), 479-488.

DISTRIBUTION LIST

Office of Naval Research

Stuart L. Brodsky
Group Leader/Mathematics
Code 411MA
Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217

Neal D. Glassman
Code 411MA
Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217

Charles J. Holland
Code 411MA
Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217

CDR Kent Hull
Code 410B
Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217

Mr. Gerald Malecki
Engineering Psychology Programs
Code 442
Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217

Charles R. Paoletti
Administrative Contracting Officer
Room 718
Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217

DCASMA, Attn: Mrs. Thelma DeMarino
Administrative Contracting Officer
300 E. Joppa Rd.
Towson, MD 21204

Dr. James C. T. Pool
Division Leader/Mathematical
& Physical Sciences
Code 411
Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217

J. Randolph Simpson (10 copies)
Scientific Officer, Code 431
Naval Analysis Programs
Mathematical and Information
Sciences Division
Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217

Dr. Martin A. Tolcott
Division Leader, Psychological
Sciences Division
Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217

Dr. Thomas C. Varley
Group Leader
Code 411OR
Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217

Edward J. Wegman
Group Leader, Code 411S&P
Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217

Office of Naval Research
Eastern/Central Regional Office
666 Summer St.
Boston, MA 02210

Other Government Agencies

Defense Technical Information
    Center (12 copies)
Cameron Station
Building 5
Alexandria, VA   22314

Naval Research Laboratory (6 copies)
Attn: Code 2627
Washington, D.C.   20375

Supplemental Distribution

Professor Kenneth J. Arrow
Department of Economics
Stanford University
Stanford, CA   94305

Michael Athens
Massachusetts Institute of Tech-
    nology
77 Massachusetts Avenue
Cambridge, MA   02139

Dr. David Castanon
Alpha-Tech
Burlington, MA   01803

Dr. Ward Edwards
Director
Social Sciences Research Institute
1042 W. 36th Street
Denney Research Building
Los Angeles, CA   90007

Supplemental Distribution (cont.)

Dr. Peter Farquhar II
Graduate School of Administration
Voorhies Hall
University of California
Davis, CA   95616

Irwin R. Goodman
Naval Ocean Systems Center
Code 7232
San Diego, CA   92152

Professor James March
Graduate School of Business
    Administration
Stanford University
Stanford, CA   94305

Dr. Amos Tversky
Department of Psychology
Stanford University
Stanford, CA   94305

Professor Robert Wilson
Department of Business Administra-
    tion
Stanford University
Stanford, CA   94305

Professor L. Zadeh
Department of Engineering
University of California
Berkeley, CA   94701

DATE
ME